

Hyperparameter Tuning

Hyperparameters:

a Arguably most important

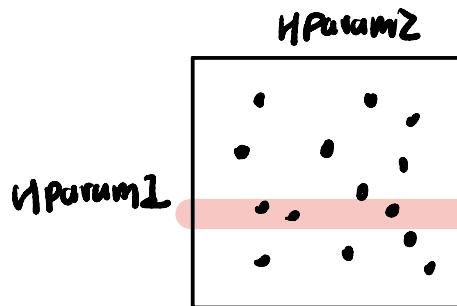
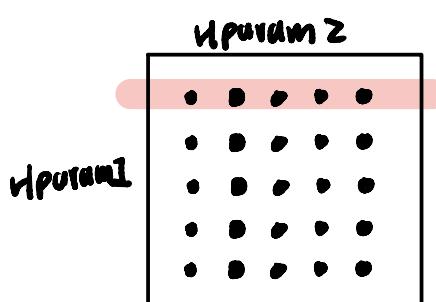
b, # hidden units, mini-batch size

layers, learning rate decay

ρ_1, ρ_2, ϵ (Adam) · Rarely tuned

....

Random Search Vs. Grid Search.



Try out more Combinations.
More effective

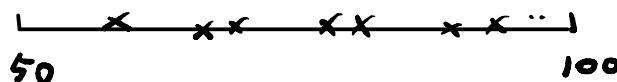
Same Hyperparam 1 for different

Hyperparam 2

Coarse to fine: Reduce searching area, perform more accurate search at more specific space.

Appropriate Scale for hyperparameters.

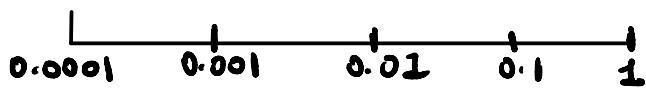
Picking hyperparameters at random:



Works for HParams such as: # layers.

Appropriate Scale (logarithmic scale).

$$r = -4 * \text{np.random.rand}()$$



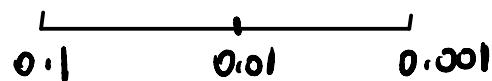
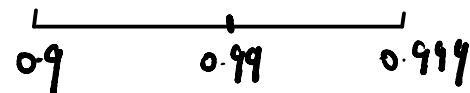
$$\alpha = 10^r$$

Hyperparameters for exponentially weighted averages.

$$\beta = 0.9 \dots 0.999$$

$$\begin{matrix} \uparrow & r \\ 10 & 1000 \end{matrix}$$

$$1-\beta = 0.1 \dots 0.001$$



$$\beta: 0.9000 \longrightarrow 0.9005 \sim 10$$

$$r \in [-3, -1]$$

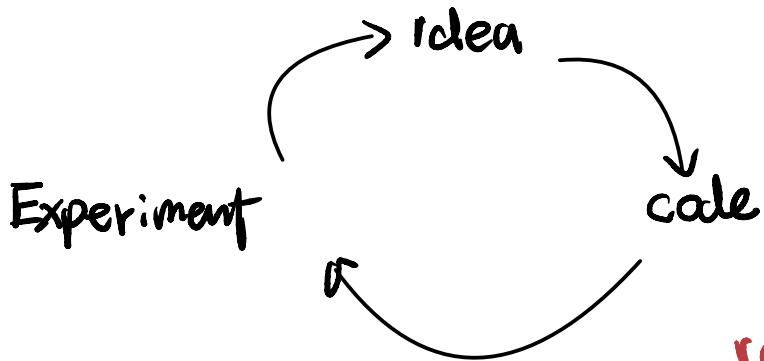
$$\beta: \frac{0.999}{1000} \longrightarrow \frac{0.9995}{2000} \sim 1000$$

$$1-\beta = 10^r$$

$$\beta = 1 - 10^r$$

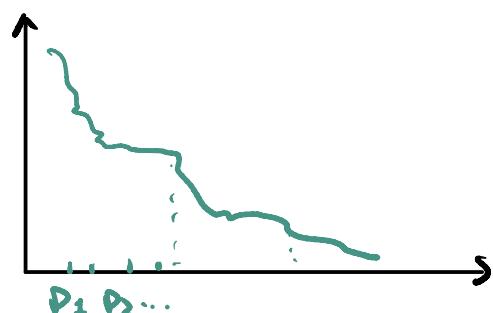
Pandas vs. Caviar

Dev process:



require intensive computation power

Pandas (Babysitting one model)



Caviar (Training many models in parallel)

